

# TTP: Tool for Tumor Progression

Johannes G. Reiter<sup>1</sup>, Ivana Bozic<sup>2,3</sup>, Krishnendu Chatterjee<sup>1</sup>, Martin A. Nowak<sup>2,3,4</sup>

<sup>1</sup> IST Austria (Institute of Science and Technology Austria), Klosterneuburg, Austria

<sup>2</sup> Program for Evolutionary Dynamics, Harvard University, Cambridge, USA

<sup>3</sup> Department of Mathematics, Harvard University, Cambridge, USA

<sup>4</sup> Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, USA

**Abstract.** In this work we present a flexible tool for tumor progression, which simulates the evolutionary dynamics of cancer. Tumor progression implements a multi-type branching process where the key parameters are the fitness landscape, the mutation rate, and the average time of cell division. The fitness of a cancer cell depends on the mutations it has accumulated. The input to our tool could be any fitness landscape, mutation rate, and cell division time, and the tool produces the growth dynamics and all relevant statistics.

## 1 Introduction

Cancer is a genetic disease which is driven by the somatic evolution of cells [1], where driver mutations for cancer increase the reproductive rate of cells through different mechanisms, e.g. evading growth suppressors, sustaining proliferative signaling, or resisting cell death [2]. Tumors are initiated by some genetic event which increases the reproductive rate of previously normal cells. The evolution of cancer (malignant tumor) is a multi-step process where cells need to receive several mutations subsequently [3]. This phase of tumor progression is characterized by the uncontrolled growth of cells [2,4]. The requirement to accumulate multiple mutations over time explains the increased risk of cancer with age.

There are several mathematical models to explain tumor progression and the age incidence of cancer [5,6,7]. The models have also provided quantitative insights in the evolution of resistance to cancer therapy [8]. The models for tumor progression are multi-type branching processes which represent an exponentially growing heterogeneous population of cells, where the key parameters for the process are: (i) the fitness landscape of the cells (which determine the reproductive rate), (ii) the mutation rate (which determines the accumulation of driver mutations), and (iii) the average cell division time (or the generation time for new cells). The fitness landscapes allow the analysis of the effects of interdependent (driver) mutations on the evolution of cancer [9].

In this work, we present a very flexible tool (namely, TTP, tool for tumor progression) to study the dynamics of tumor progression. The input to our tool could be any fitness landscape, mutation rate, and cell division time, and the tool generates the growth dynamics and all relevant statistics (such as the expected

tumor detection time, or the expected appearance time of surviving mutants, etc). Our stochastic computer simulation is an efficient simulation of a multi-type branching process under all possible fitness landscapes, driver mutation rates, and cell division times.

Our tool provides a quantitative framework to study the dynamics of tumor progression in different stages of tumor growth. Currently, the data to understand the effects of complex fitness landscapes can only be obtained from patients or animals suffering the disease. With our tool, playing with the parameters, once the real-world data is reproduced, the computer simulations can provide many simulation examples that would aid to understand these complex effects. Moreover, once the correct mathematical models for specific types of cancer are identified where the simulations match the real-world data, verification tools for probabilistic systems can be used to further analyze and understand the tumor progression process (such an approach has been followed in [10] for the verification of biological models). In this direction, results of specific fitness landscapes of our tool have already been used in a biological application paper [11]. While we present our tool for the discrete-time process (which provides a good approximation of the continuous-time process), results of our tool for the special case of a uniform fitness landscape in the continuous-time process have also been shown to have excellent agreement with the real-world data for the time to treatment failure for colorectal cancer [8].

## 2 Model

Tumor progression is modeled as a discrete-time branching process (Galton-Watson process [12]). At each time step, a cell can either divide or die. The phenotype  $i$  of a cancerous cell determines its division probability  $b_i$  and is encoded as a bit string of length four (i.e.  $\{0,1\}^4$ ). The death probability  $d_i$  follows from  $b_i$  as  $d_i = 1 - b_i$ . If a cell divides, one of the two daughter cells can receive an additional mutation (i.e., a bit flips from wildtype 0 to the mutated type 1) with probability  $u$  in one of the wildtype positions (e.g., cells of phenotype 1010 can receive an additional mutation only at positions two and four; cells of phenotype 1111 can not receive any additional mutations). The branching process is initiated by a single cell of phenotype  $i = 0000$  (resident cell). The resident cells are wildtype at all four positions and have a strictly positive growth rate (i.e.,  $b_{0000} - d_{0000} > 0$ ).

**Fitness landscapes.** Our tool provides two predefined fitness landscapes for driver mutations in tumor progression: (1) Multiplicative Fitness Landscape (MFL) and (2) Path Fitness Landscape (PFL). Additionally, the user can also define its own general fitness landscape (GFL). A fitness landscape defines the birth probability  $b_i$  for all possible phenotypes  $i$ . Following the convention of the standard modeling approaches, we let  $b_{0000} = 1/2(1 + s_0)$  be the birth probability of the resident cells (i.e., cells of phenotype 0000) [9,11]. The growth coefficient  $s_j$  indicates the selective advantage provided by an additional mutation at position  $j$  in the phenotype.

*Multiplicative fitness landscape.* In the MFL a mutation at position  $j$  of the phenotype  $i$  of a cell results in a multiplication of its birth probability by  $(1+s_j)$ . Specifically, the birth probability  $b_i$  of a cell with phenotype  $i$  is given by

$$b_i = \frac{1}{2}(1+s_0) \prod_{j=1}^4 (1+\hat{s}_j);$$

where  $\hat{s}_j = 0$  if the  $j$ -th position of  $i$  is 0; otherwise  $\hat{s}_j = s_j$ . Hence, each additional mutation can be weighted differently and provides a predefined effect  $(1+s_1)$ ,  $(1+s_2)$ ,  $(1+s_3)$ , or  $(1+s_4)$  on the birth probability of a cell. Additional mutations can also be costly (or neutral) which can be modeled by a negative  $s_j$  (or  $s_j = 0$ ). If  $s_0 = s_1 = s_2 = s_3 = s_4$  the fitness landscape reduces to the model studied by Bozic et al. [9], which we call EMFL (Equal Multiplicative Fitness Landscape) and is also predefined in our tool.

*Path fitness landscape.* The PFL defines a certain path on which additional mutations need to occur to increase the birth probability of a cell. The predefined path can be  $0000 \rightarrow 1000 \rightarrow 1100 \rightarrow 1110 \rightarrow 1111$ , and again the growth coefficients  $s_j$  determine the multiplicative effect of the new mutation on the birth probability (see appendix for more details). Mutations not on this path are deleterious for the growth rate of a cell and its birth probability is set to  $1/2(1-v)$ . The parameter  $v$  ( $0 \leq v \leq 1$ ) specifies the disadvantage for cells of all phenotypes which do not belong to the given path.

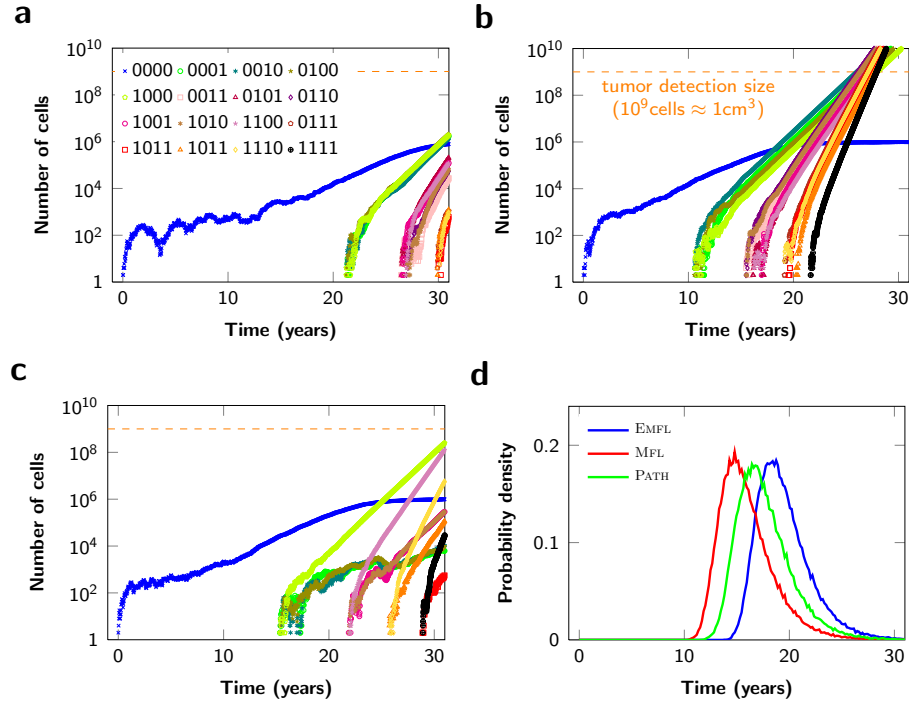
*General fitness landscapes.* Our tool allows to input any fitness landscape as follows: for  $b_i$  for  $i \in \{0,1\}^4$ , our tool can take as input the value of  $b_i$ . In this way, any fitness landscape can be a parameter to the tool.

**Density limitation.** In some situations, a tumor needs to overcome current geometric or metabolic constraints (e.g. when the tumor needs to develop blood vessels to provide enough oxygen and nutrients for further growth [13,11]). Such growth limitations are modeled by a density limit (carrying capacity) for various phenotypes. Hence, the cells of a phenotype  $i$  grow first exponentially but eventually reach a steady state around a given carrying capacity  $K_i$ . Only cells with another phenotype (additional mutation) can overcome the density limit. Logistic growth is modeled with variable growth coefficients  $\tilde{s}_j = s_j(1 - X_i/K_i)$  where  $X_i$  is the current number of cells of phenotype  $i$  in the tumor. In this model, initially  $\tilde{s}_j \approx s_j$  ( $X_i \ll K_i$ ), however, if  $X_i$  is on the order of  $K_i$ ,  $\tilde{s}_j$  becomes approximately zero (details are given in the appendix).

### 3 Tool Implementation & Experimental Results

Our tool provides an efficient implementation of a very general tumor progression model. Essentially, the tool implements the above defined branching processes to simulate the dynamics of tumor growth and to obtain statistics about the expected tumor detection time and the appearance of additional driver mutations during different stages of disease progression. TTP can be downloaded from here: <http://pub.ist.ac.at/ttp>.

For an efficient processing of the discrete-time branching process, the stochastic simulation samples from a multinomial distribution for each phenotype at each time step [9,11]. The sample returns the number of cells which divided with and without mutation and the number of cells which died in the current generation (see the appendix for details). From the samples for each phenotype the program calculates the phenotype distribution in the next generation. Hence, the program needs to store only the number of cells of each phenotype during the simulation. This efficient implementation of the branching process allows the tool to simulate many patients within a second and to obtain very good statistical results in a reasonable time frame.



**Fig. 1.** Experimental results illustrating the variability of tumor progression. In panels a and b we show examples for two particular simulation runs where the cells grow according to the EMFL and resident cells (blue) are constrained by a carrying capacity of  $K_{0000} = 10^6$ . In panel c the cells grow according to the PFL. In panel d we show statistical results for the probability density of tumor detection when cells grow according to different fitness landscapes. Parameter values: growth coefficients  $s_0 = 0.004$ ,  $s_1 = 0.006$ ,  $s_2 = 0.008$ ,  $s_3 = 0.01$ ,  $s_4 = 0.012$ , and  $v = 0.01$ , mutation rate  $u = 10^{-6}$ , cell division time  $T = 3$  days, tumor detection size  $10^9$  cells.

**Modes.** The tool can run in the following two modes: individual or statistics. In the *individual mode* the tool produces the growth dynamics of one tumor in a patient (see panels a, b, and c in Fig. 1). Furthermore, both the growth dynamics and the phenotype distribution of the tumor are depicted graphically. In the *statistics mode* the tool produces the probability distribution for the detection time of the tumor (see panel d in Fig. 1) both graphically and quantitatively. Additionally, the tool calculates for all phenotypes the appearance times of the first surviving lineage, the existence probability, and the average number of cells at detection time.

**Features.** TTP provides an intuitive graphical user interface to enter the parameters of the model and shows plots of the dynamics during tumor progression, the phenotype distribution, or the probability density of tumor detection. These plots can also be saved as files in various image formats. Furthermore, the tool can create data files (tab-separated values) of the tumor growth history and the probability distribution of tumor detection for any set of input parameters (details on the format are given in the appendix).

**Input parameters.** In both modes, the tool takes the following input parameters: (i) growth coefficients  $s_0$ ,  $s_1$ ,  $s_2$ ,  $s_3$ , and  $s_4$  (and  $v$  in the case of PFL), (ii) mutation rate  $u$ , (iii) cell generation time  $T$ , (iv) fitness landscape (MFL, PFL, EMFL, or GFL with the birth probability for each phenotype), and optional (v) density limits for some phenotypes. In the individual mode, additionally, the user needs to provide the number of generations which have to be simulated. In the statistics mode, the additional parameters are: the tumor detection size and the number of patients (tumors which survive the initial stochastic fluctuations) which have to be simulated.

**Experimental results.** In panels a, b, and c of Fig. 1 we show examples of the growth dynamics during tumor progression. Although we used exactly the same parameters in panels a and b, we observe that the time from tumor initiation until detection can be very different. In panel d we show the probability density of tumor detection under various fitness landscapes. Further experimental results are given in the appendix.

**Case studies.** Several results of these models have shown excellent agreement with different aspects of real-world data. In [9], results for the expected tumor size at detection time using a EMFL fit the reported polyp sizes of the patients very well. Similarly, using a continuous-time branching process and a uniform fitness landscape, results for the expected time to the relapse of a tumor after start of treatment agree thoroughly with the observed times in patients [8].

**Future work.** In some ongoing work, we also investigate mathematical models for tumor dynamics occurring during cancer treatment modeled by a continuous-time branching process. Thus an interesting extension of our tool would be to model treatment as well. Another interesting direction is to model the seeding of metastasis during tumor progression and hence simulate a “full” patient rather than the primary tumor alone. Once faithful models of the evolution of cancer have been identified, verification tools such as PRISM [14] and theoretical results such as [15] might contribute to the understanding of these processes.

**Acknowledgments.** This work is supported by the ERC Start grant (279307: Graph Games), the FWF NFN Grant (No S11407-N23, Rise), the FWF Grant (No P 23499-N23), a Microsoft Faculty Fellow Award, the Foundational Questions in Evolutionary Biology initiative of the John Templeton Foundation, and the NSF/NIH joint program in mathematical biology (NIH grant R01GM078986).

## References

1. Vogelstein, B., Kinzler, K.W.: Cancer genes and the pathways they control. *Nature medicine* **10**(8) (2004) 789–799
2. Hanahan, D., Weinberg, R.a.: Hallmarks of cancer: the next generation. *Cell* **144**(5) (2011) 646–674
3. Jones, S., Chen, W.D., Parmigiani, G., Diehl, F., Beerenwinkel, N., Antal, T., Traulsen, A., Nowak, M.A., Siegel, C., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Willis, J., Markowitz, S.D.: Comparative lesion sequencing provides insights into tumor evolution. *PNAS* **105**(11) (2008) 4283–4288
4. Nowak, M.A.: *Evolutionary Dynamics: Exploring the equations of life*. The Belknap Press of Harvard University Press, Cambridge, MA (2006)
5. Komarova, N.L., Sengupta, A., Nowak, M.A.: Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of Theoretical Biology* **223**(4) (2003) 433–450
6. Iwasa, Y., Michor, F., Nowak, M.A.: Stochastic tunnels in evolutionary dynamics. *Genetics* **166**(3) (2004) 1571–1579
7. Nowak, M.A., Michor, F., Komarova, N.L., Iwasa, Y.: Evolutionary dynamics of tumor suppressor gene inactivation. *PNAS* **101**(29) (2004) 10635–10638
8. Diaz, L.A., Williams, R.T., Wu, J., Kinde, I., Hecht, J.R., Berlin, J., Allen, B., Bozic, I., Reiter, J.G., Nowak, M.A., Kinzler, K.W., Oliner, K.S., Vogelstein, B.: The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**(7404) (2012) 537–540
9. Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., Nowak, M.A.: Accumulation of driver and passenger mutations during tumor progression. *PNAS* **107**(43) (2010) 18545–18550
10. Sadot, A., Fisher, J., Barak, D., Admanit, Y., Stern, M., Hubbard, E., Harel, D.: Toward verified biological models. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* **5**(2) (2008) 223–234
11. Reiter, J.G., Bozic, I., Allen, B., Chatterjee, K., Nowak, M.A.: The effect of one additional driver mutation on tumor progression. *Evolutionary Applications* **6**(1) (2012) 34–45
12. Haccou, P., Jagers, P., Vatutin, V.A.: *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press (2005)
13. Kerbel, R.S.: Tumor angiogenesis: past, present and the near future. *Carcinogenesis* **21**(3) (March 2000) 505–515
14. Hinton, A., Kwiatkowska, M.Z., Norman, G., Parker, D.: Prism: A tool for automatic verification of probabilistic systems. In: *TACAS*. (2006) 441–444
15. Etessami, K., Stewart, A., Yannakakis, M.: Polynomial time algorithms for multi-type branching processes and stochastic context-free grammars. In: *STOC*. (2012) 579–588

## A Appendix: Details of the Tool

TTP is available for download at <http://pub.ist.ac.at/ttp>. The tool is implemented in Java and runs on all operating systems which run a Java Virtual Machine (JVM) of version 1.7 or above ([www.java.com/getjava](http://www.java.com/getjava)). All the necessary libraries are included in the jar-file.

### A.1 Tool Features

Our tool supports various features in two running modes.

In the individual mode, TTP simulates the tumor growth dynamics for a given number of generations. Plots of the growth dynamics over time and the current phenotype distribution are produced simultaneously. (Both plots can be saved in a PNG-file or SVG-file.) The full growth history for all cell types can also be stored in a data-file (format is described in section A.2).

In the statistics mode, TTP simulates the given number of patients with the same parameters and simultaneously shows the probability density of tumor detection for a given detection size ( $10^9$  cells correspond to a tumor volume of approximately  $1\text{cm}^3$ ). The average tumor detection time and the average fraction of resident cells at detection are also shown during the simulations. After all patients have been simulated, the existence probability at detection, the average number of cells, and the average appearance year of the first surviving cell for all phenotypes are calculated and shown in a new window. In addition the tool shows the number of detected and died tumors per year in a separate window. All these data is stored in a data-file (format is described in section A.2).

### A.2 Installation and Implementation Details

TTP is written in Java and makes use of several other libraries. The tool requires the Java Runtime Environment (JRE) of version 1.7 or above. To start TTP double-click on `ttp.jar` or on the command line type: `java -jar ttp.jar`. (Make sure that you have the permission to execute `ttp.jar`. On Mac OS invoking the tool from the command line can overcome the security restrictions.)

The tool is composed of the following components: the model implementation, the statistics thread, the graphical user interface, and the plot generator.

**Model implementation.** The core component of the tool is the efficient implementation of the discrete-time branching process. Following Bozic et al. [9], the number of cells  $X_i$  of phenotype  $i$  in the next generation ( $t+1$ ) is calculated by sampling from the multinomial distribution

$$\text{Prob}[(Y_1, Y_2, Y_3) = (y_1, y_2, y_3)] = \frac{X_i(t)!}{y_1!y_2!y_3!} [b_i(1-u)^{y_1} d_i^{y_2} (b_i u)^{y_3}] \quad (1)$$

where  $y_1 + y_2 + y_3 = X_i(t)$ ,  $y_3 = \sum_k M_{ik}$ , and

$$X_i(t+1) = X_i(t) + Y_1 - Y_2 + \sum_k M_{ki} . \quad (2)$$

The number of cells which give birth to an identical daughter cell is denoted by  $Y_1$ , the number of cells which die is denoted by  $Y_2$ . The number of cells which divide with an additional mutation is given by  $Y_3$  and the number of cells mutated from phenotype  $k$  to  $i$  is given by  $M_{ki}$ . In general, one can define a mutation matrix to encode the probabilities  $m_{ki}$  that a cell of phenotype  $k$  mutates to a cell of phenotype  $i$ . In our case this matrix is defined by the sequential accumulation of mutations. A cell of some phenotype can receive an additional mutation only on the positions in its bit-string encoding which are wildtype (i.e. only bit flips from 0 to 1 are allowed). Mutations on all allowed positions are equally likely. Back mutations are not considered.

*Fitness landscapes.* Our tool supports four fitness landscapes for additional driver and passenger mutations: (i) MFL, (ii) EMFL, (iii) PFL, and (iv) GFL. In principal, driver mutations increase the birth rate of a cell whereas passenger mutations have no effect on the cell's birth rate [9]. In the tables A1 and A2 we present the complete definition of the MFL and the PFL, respectively. (The definition of EMFL and GFL have been given in Section 2.)

**Table A1.** Multiplicative Fitness Landscape.

Additional mutations	Phenotype	Birth probability
0	0000	$1/2(1 + s_0)$
1	1000	$1/2(1 + s_0)(1 + s_1)$
	0100	$1/2(1 + s_0)(1 + s_2)$
	0010	$1/2(1 + s_0)(1 + s_3)$
	0001	$1/2(1 + s_0)(1 + s_4)$
2	1100	$1/2(1 + s_0)(1 + s_1)(1 + s_2)$
	1010	$1/2(1 + s_0)(1 + s_1)(1 + s_3)$
	1001	$1/2(1 + s_0)(1 + s_1)(1 + s_4)$
	0110	$1/2(1 + s_0)(1 + s_2)(1 + s_3)$
	0101	$1/2(1 + s_0)(1 + s_2)(1 + s_4)$
	0011	$1/2(1 + s_0)(1 + s_3)(1 + s_4)$
3	1110	$1/2(1 + s_0)(1 + s_1)(1 + s_2)(1 + s_3)$
	1101	$1/2(1 + s_0)(1 + s_1)(1 + s_2)(1 + s_4)$
	1011	$1/2(1 + s_0)(1 + s_1)(1 + s_3)(1 + s_4)$
	0111	$1/2(1 + s_0)(1 + s_1)(1 + s_3)(1 + s_4)$
4	1111	$1/2(1 + s)(1 + s_1)(1 + s_2)(1 + s_3)(1 + s_4)$

*Density limit.* Our tool allows a separate carrying capacity  $K_i$  for each phenotype  $i$ . When the GFL is used, in the beginning of the simulation, the growth coefficients  $s_i$  are calculated from the given  $b_i$  for all phenotypes  $i \in \{0, 1\}^4$  since the density limiting effects are based on the values for  $\tilde{s}_i$ . As a technical detail, for sizes  $X_i$  for which the birth probability  $b_i$  would fall below 0 (or for which, equivalently,  $\tilde{s}_i$  would fall below -1), we set  $b_i = 0$ .

**Statistics thread.** The statistics thread handles the simulation of many identical branching processes to obtain the statistical results. These simulations run



**Table A2.** Path Fitness Landscape.

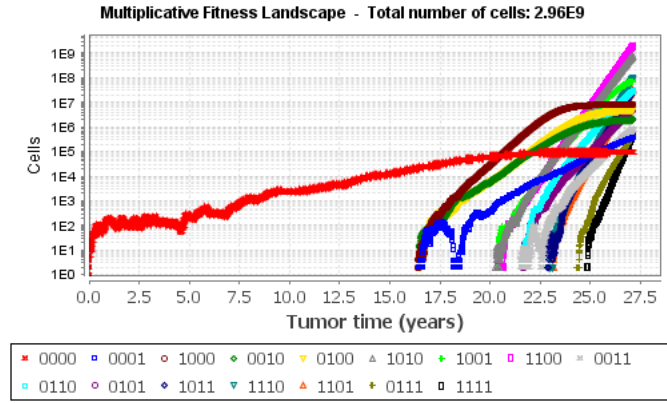
Additional mutations	Phenotype	Birth probability
0	0000	$\frac{1}{2}(1 + s)$
	1000	$\frac{1}{2}(1 + s_0)(1 + s_1)$
1	0100	$\frac{1}{2}(1 - v)$
	0010	$\frac{1}{2}(1 - v)$
	0001	$\frac{1}{2}(1 - v)$
	1100	$\frac{1}{2}(1 + s_0)(1 + s_1)(1 + s_2)$
2	1010	$\frac{1}{2}(1 - v)$
	1001	$\frac{1}{2}(1 - v)$
	0110	$\frac{1}{2}(1 - v)$
	0101	$\frac{1}{2}(1 - v)$
	0011	$\frac{1}{2}(1 - v)$
	1110	$\frac{1}{2}(1 + s_0)(1 + s_1)(1 + s_2)(1 + s_3)$
3	1101	$\frac{1}{2}(1 - v)$
	1011	$\frac{1}{2}(1 - v)$
	0111	$\frac{1}{2}(1 - v)$
4	1111	$\frac{1}{2}(1 + s_0)(1 + s_1)(1 + s_2)(1 + s_3)(1 + s_4)$

in a separate thread such that the GUI keeps responsive for user requests. After completing all the necessary simulations a data-file with all relevant results is automatically generated and stored to the execution directory of the tool.

**Graphical user interface.** The graphical user interface (GUI) component contains frames and forms required for the functionality of the tool. It also handles all the user requests and distributes them to the other components. Within the GUI the plots for the tumor progression dynamics (in individual mode) and for the probability density of tumor detection (in statistics mode) are displayed. Multiple screenshots of the GUI are shown in section A.4.

**Plot generator.** The plot generation is based on the free JFreeChart library ([www.jfree.org](http://www.jfree.org)). For the generation of the Scalable Vector Graphics (SVG) the Apache XML Graphics library (<http://xmlgraphics.apache.org>) is used. An example for a plot generated by our tool is shown in Figure A2.

**Data files.** TTP produces various data files which can be used for further analysis and processing. The data are given as tab-separated values where each record is one line of the text files. In Listing A1 we show an example of a data file generated in the statistics mode. Average results are given as comments which start with a hash. In the individual mode the data file contains the number of cells  $X_i$  for each phenotype  $i$  in all generations.



**Fig. A2.** Example for a generated plot of the tumor growth dynamics.

**Listing A1.** Generated data file in the statistics mode.

```

# Used fitness landscape: Multiplicative
# Growth coefficients: s0=0.004, s1=0.006, s2=0.008, s3=0.01, s4=0.012
# Mutation rate: 1E-6
# Generation time (days): 3.0
# Simulation for 10000 patients.
# Bin size is 20 generations.
generation detected      cumul_det      died      cumul_died
0            0            0            0.046003   0.920069
20           0            0            0.001938   0.958826
40           0            0            0.000705   0.972923
60           0            0            0.000358   0.98009
80           0            0            0.000227   0.984622
100          0            0            0.000149   0.987594
.            .            .            .            .
.            .            .            .            .
5980         0            1            0            1
6000         0            1            0            1
# 0.7981% (abs: 10000) have been detected within at most
#      20000 generations.
# 99.2% (abs: 1243006) tumors went extinct.
# Average year of detection: 22.04
# Mutant appearance times:
# Mutant 0000 appeared in average in generation: 0 (year: 0),
#      existance probability: 100%,
#      average number of cells: 3267013.74
# Mutant 0001 appeared in average in generation: 1540 (year: 12.66),
#      existance probability: 100%,
#      average number of cells: 1925660.02
.            .            .            .            .
.            .            .            .            .
# 1253006 runs have been performed to create 10000 patients.

```

### A.3 User Manual

TTP is invoked by a double-click on `ttp.jar` or by the command `java -jar ttp.jar`. After the tool has started, the GUI can be used for all operations (see Figure A3 for a screenshot of the GUI).

**Input parameters.** In the control panel the tool takes all the main parameters for tumor progression: the fitness landscape, the mutation rate, and the cell division time. If one of the prespecified fitness landscapes (MFL, PFL, or EMFL) is used, the relevant growth coefficients have to be defined. If the general fitness landscape is used, a pop-up window appears after the selection of the GFL and the specific birth probabilities for all the phenotypes can be defined. To add density limits for specific phenotypes, a pop-up window appears after “density limit” has been checked. For each phenotype a different density limit can be given ( $-1$  indicates that there is no limitation on this phenotype). To obtain statistical results, the number of patients (i.e. the number of tumors with a surviving lineage) and the tumor detection size (i.e. the number of cells when a tumor can be detected) need to be provided.

**Modes.** After all the parameter values have been specified, the tool can either run in the individual or the statistics mode. To simulate the growth dynamics of a single tumor, click on “New simulation” and the tool runs in the individual mode. Then any number of cell generations can be simulated until the tumor consists of more than  $10^{15}$  cells.

The statistics mode can be started by clicking on “Obtain statistics”. The tool simulates the given number of tumors until they reach the detection size and calculates all relevant statistics.

**Output.** In the individual mode, the tool generates plots for the growth dynamics and the phenotype distribution during the simulation. Furthermore, the entire tumor growth dynamics for each phenotype can be stored as a data file and the plots can be saved as PNG and SVG files (plots are stored to the folder “charts” in the execution directory of the tool).

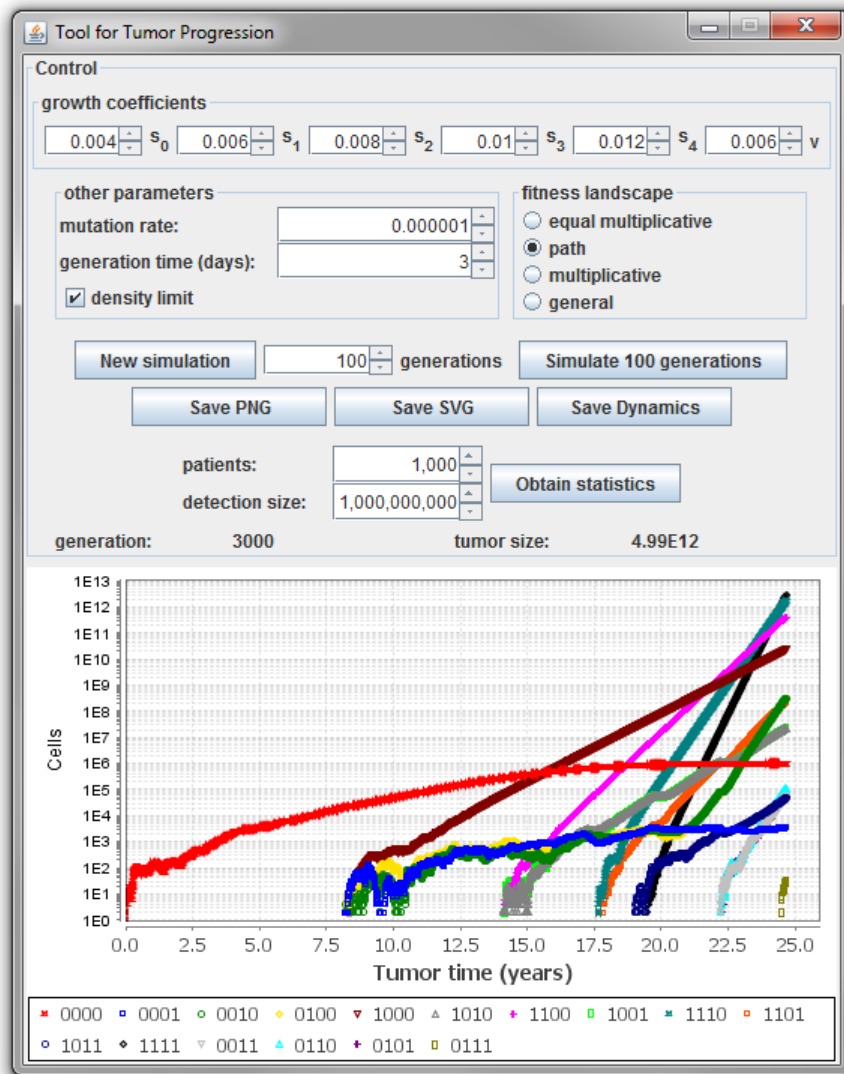
In the statistics mode, the tool generates the plot for the probability density of tumor detection. Statistics about the appearance time of the mutants and the detection and extinction year are shown in separate windows (see Figures A4 and A6 for screenshots). All the generated statistics are automatically saved to a data file (see Listing A1).

### A.4 Experimental Results & Screenshots

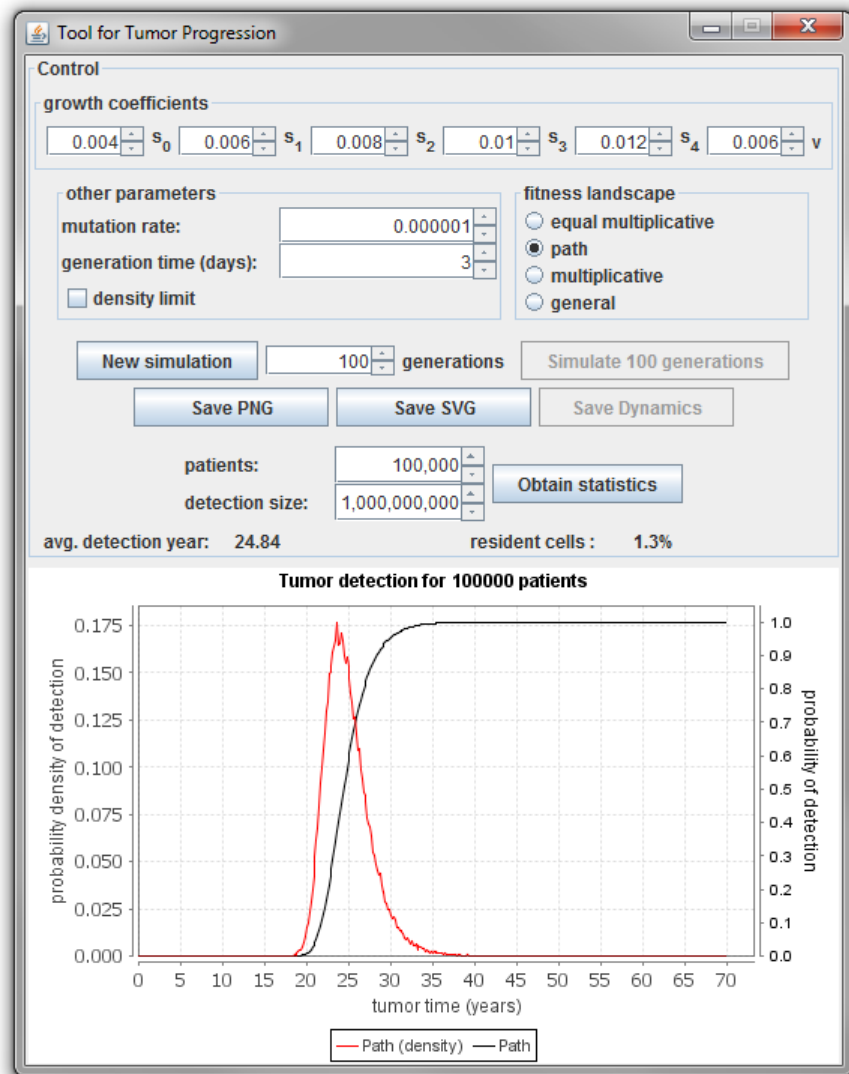
In this section we present some additional experimental results and multiple screenshots of the tool. In Table A3 we compare the probability of tumor detection for the fitness landscapes EMFL, MFL, and PFL. In average our tool needs approximately 90ms to simulate a tumor with  $10^9$  cells (on a dual core 2.67GHz processor).

**Table A3.** Cumulative probability of tumor detection for different fitness landscapes. Results are averages over  $10^5$  runs. Parameter values: growth coefficients  $s_0 = 0.004$ ,  $s_1 = 0.006$ ,  $s_2 = 0.008$ ,  $s_3 = 0.01$ ,  $s_4 = 0.012$ ,  $v = 0.006$ , mutation rate  $u = 10^{-6}$ , detection size  $M = 10^9$ .

generation	EMFL	MFL	PFL
2000	0.0	0.001	0.0
2200	0.0	0.038	0.0
2400	0.0	0.212	0.006
2600	0.004	0.491	0.075
2800	0.068	0.724	0.288
3000	0.289	0.865	0.559
3200	0.567	0.937	0.766
3400	0.774	0.971	0.887
3600	0.892	0.987	0.947
3800	0.949	0.994	0.976
4000	0.977	0.998	0.989
4200	0.989	0.999	0.995
4400	0.995	1.0	0.998
4600	0.998	1.0	0.999
4800	0.999	1.0	1.0
5000	1.0	1.0	1.0



**Fig. A3.** Graphical User Interface of TTP in the individual mode.



**Fig. A4.** Graphical User Interface of TTP in the statistics mode.

year	detected	died off
1	0	12216924
2	0	91220
3	0	28700
4	0	12818
5	0	6803
6	0	3755
7	0	2147
8	0	1285
9	0	801
10	0	463
11	0	299
12	0	176
13	0	79
14	0	60
15	0	41
16	0	24
17	0	14
18	2	8
19	73	6
20	647	0
21	3100	1
22	8181	2
23	14033	0
24	16892	0
25	16035	0
26	12955	0
27	9491	0
28	6520	0
29	4519	0
30	2898	0
31	1720	0
32	1104	0
33	670	0

Probability of tumor survival: 0.8%

**Fig. A5.** Statistical results of the average detection and extinction year.

Mutant statistics at detection time			
mutant type	appearance year	existence probability	number of cells
0000		100%	1.3E7
0001	12.8301	100%	4.05E4
0010	12.8384	100%	6.33E4
0100	12.8301	100%	6.33E4
1000	12.6247	100%	3.11E8
0011	23.9753	91.478%	115.3
0101	23.9836	91.527%	114.8
0110	23.9507	94.389%	119.9
1001	17.737	100%	3.27E5
1010	17.737	100%	3.27E5
1100	17.6384	100%	5.59E8
0111		0%	0
1011	22.9808	100%	877.6
1101	20.7041	100%	3.3E5
1110	20.663	100%	1.26E8
1111	22.7589	100%	1.81E6
Avg. detection year: 24.83 (Averages of 100000 detected tumors)			

**Fig. A6.** Statistical results of the average appearance year, the existence probability and the number of cells at detection time.